Data Mining for network topology and traffic

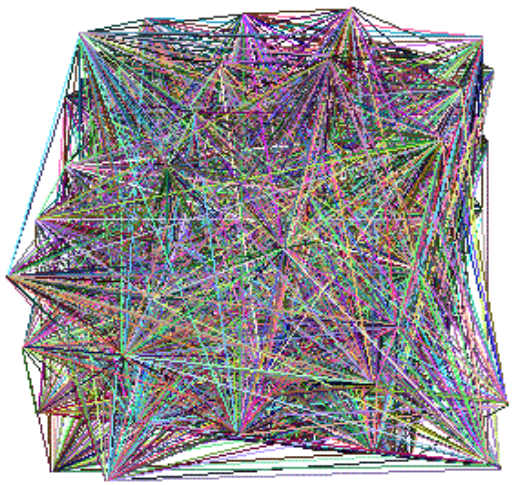
Christos Faloutsos

Carnegie Mellon University
www.cs.cmu.edu/~christos

Outline

- Fast estimation of the neighborhood function [w/ C. Palmer, M. Faloutsos, G. Siganos]
- Automatic traffic mining

Power laws in networks



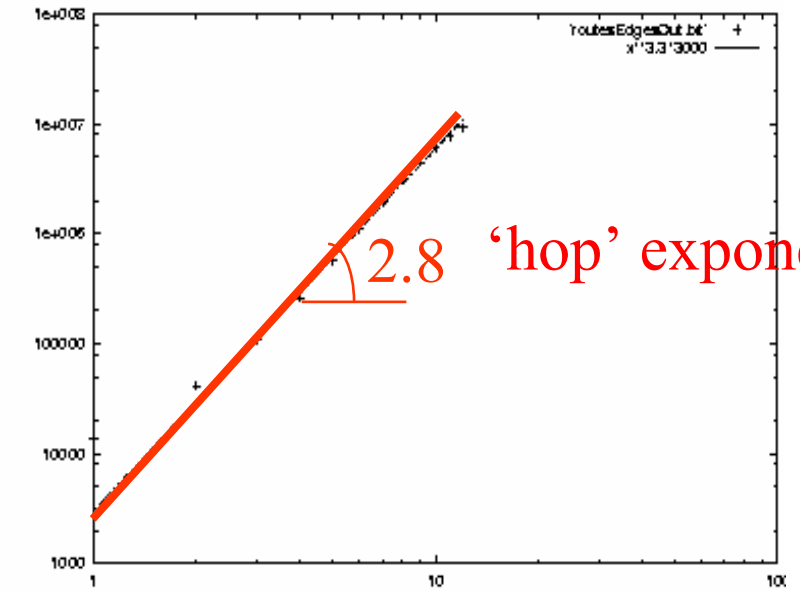
- Internet; web; gnutella P2P networks
- Q: Any pattern?
- A: power laws!

Hop Exponent *H*

- A: neighborhood function *N(h)* = number of pairs within *h* hops or less [Nicol] - power law, too!

log(#pairs)

internet

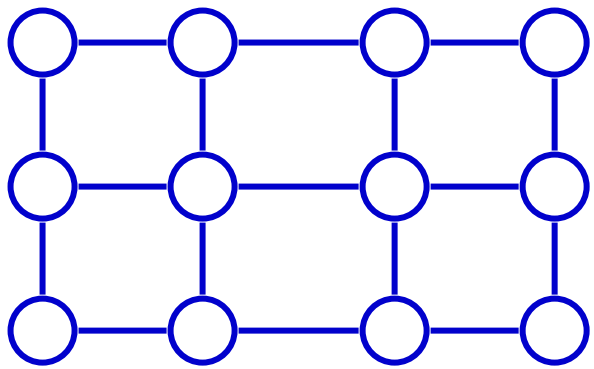


'hop' exponent

log(hops)

Hop exp. = 1

Hop exp. = 2



More on the hop exponent

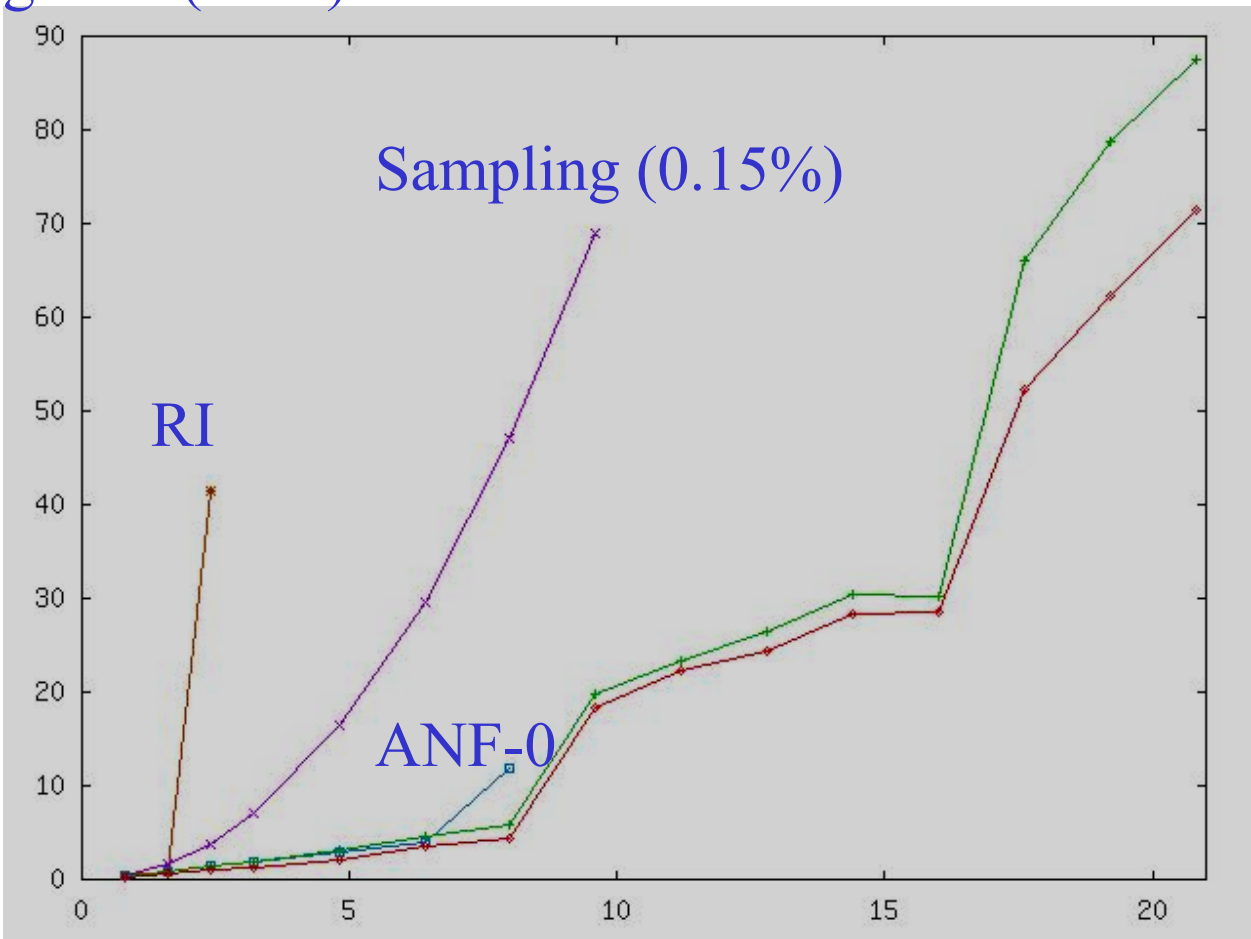
- 'Intrinsic'/fractal dimensionality of the nodes of the graph
- But: naively it needs O(N**2) (terrible for large graphs)
- What to do?

Solution:

- A: Approximation: 'ANF' (approx. neighborhood function [KDD02] - response time: from **day** to **minutes**

Scalability of ANF!

Running time (mins)



(Approx.) neighborhood function

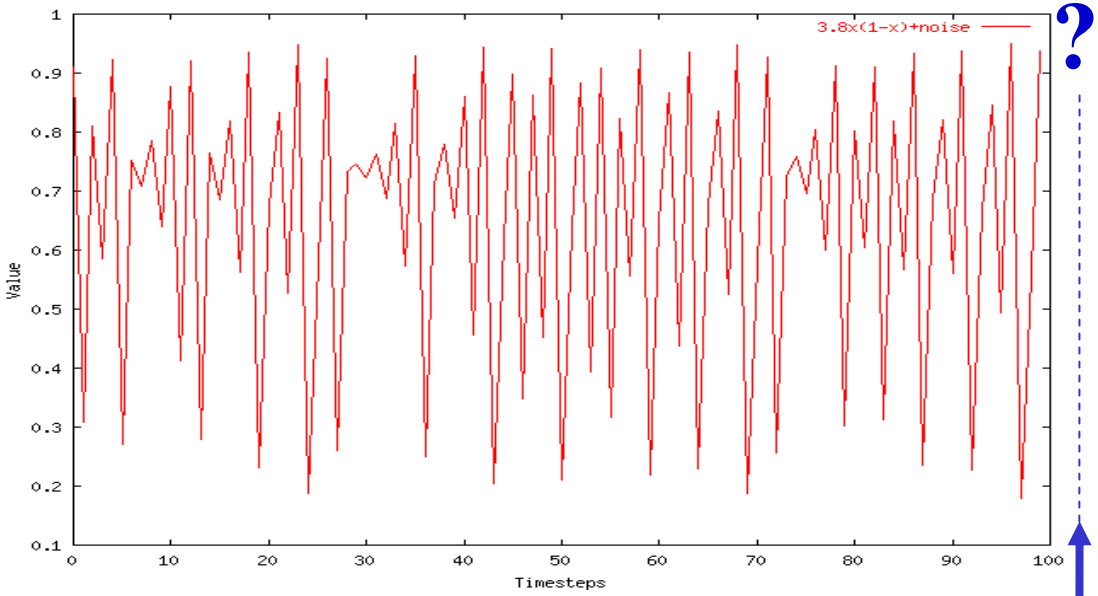
- Useful for estimating the diameter of a graph;
- the ``effective radius'' of a node (distance to 90%-tile of the other nodes)
- the connectivity under failures [Nicol]
- quick checks for (dis-)similarity between two graphs

Outline

- Fast estimation of the neighborhood function [w/ C. Palmer, M. Faloutsos, G. Siganos]
- Automatic traffic mining

Problem #2: Forecasting & Mining

Value



?

Time

Problem definition

- Given: one or more sequences

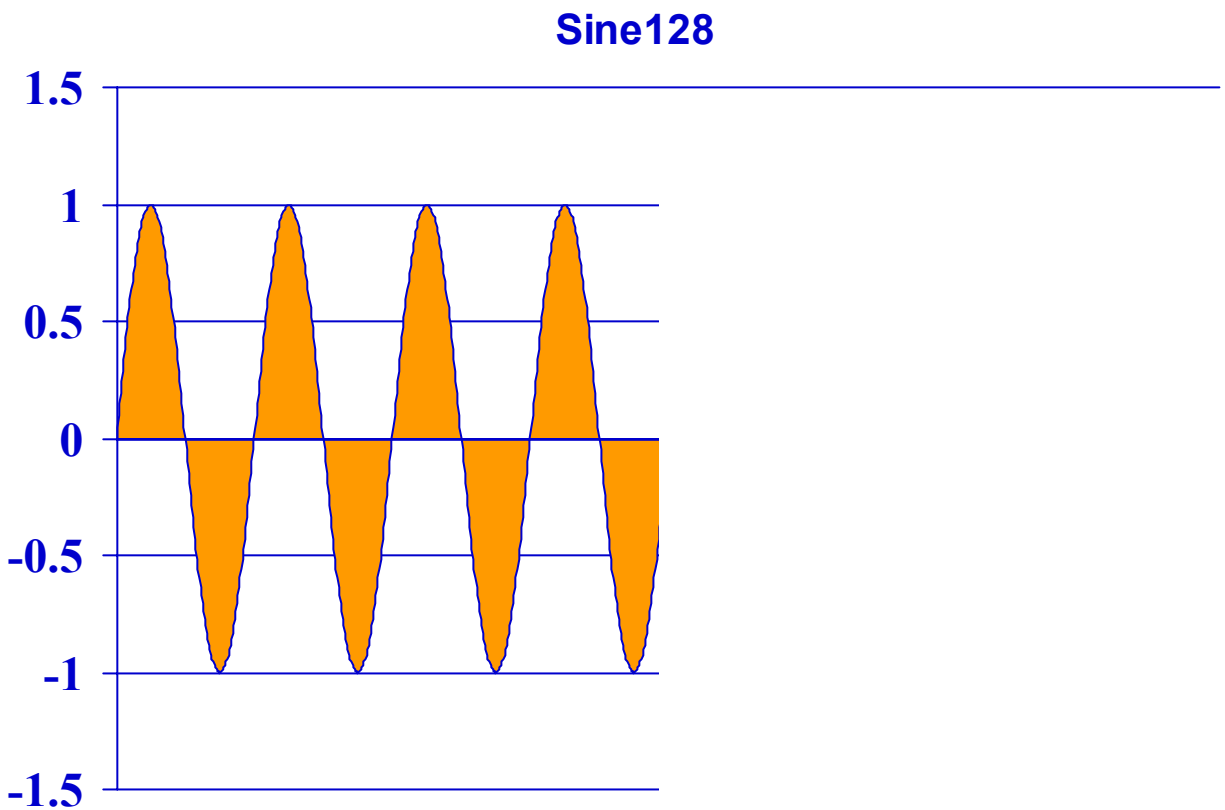
  $x_1, x_2, \ldots, x_t, \ldots; (y_1, y_2, \ldots, y_t, \ldots)$

- Find
  - forecasts; patterns; clusters; outliers
- **automatically**;
- w/ **single-pass**, **any-time algo**

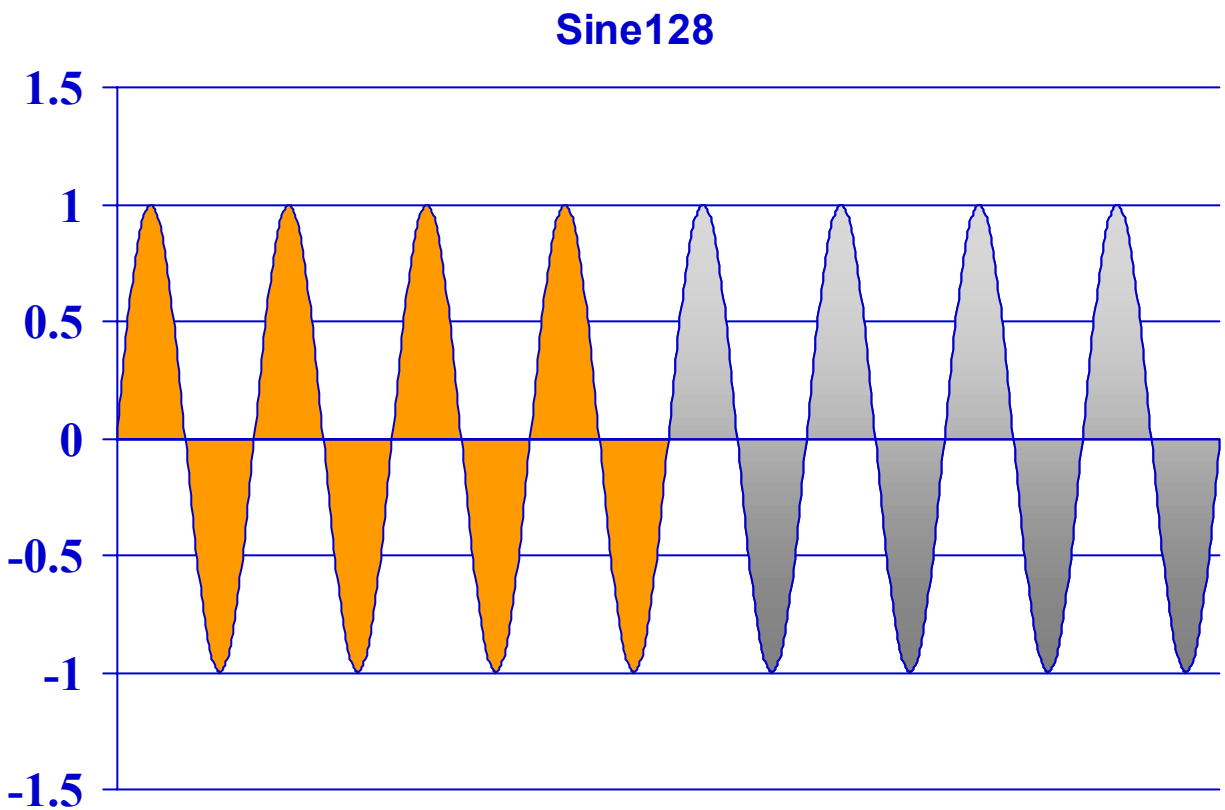
Motivation - Applications

- Network traffic modeling, AND
- Financial, sales, economic series
- Medical
- scientific/environmental
- military; industrial; ...

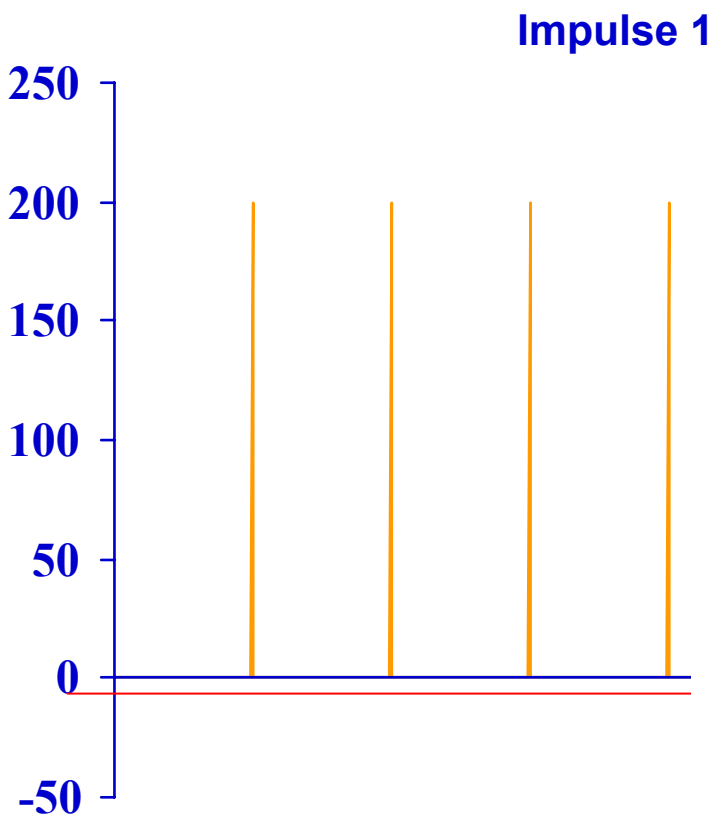
AWSOM: Some Results



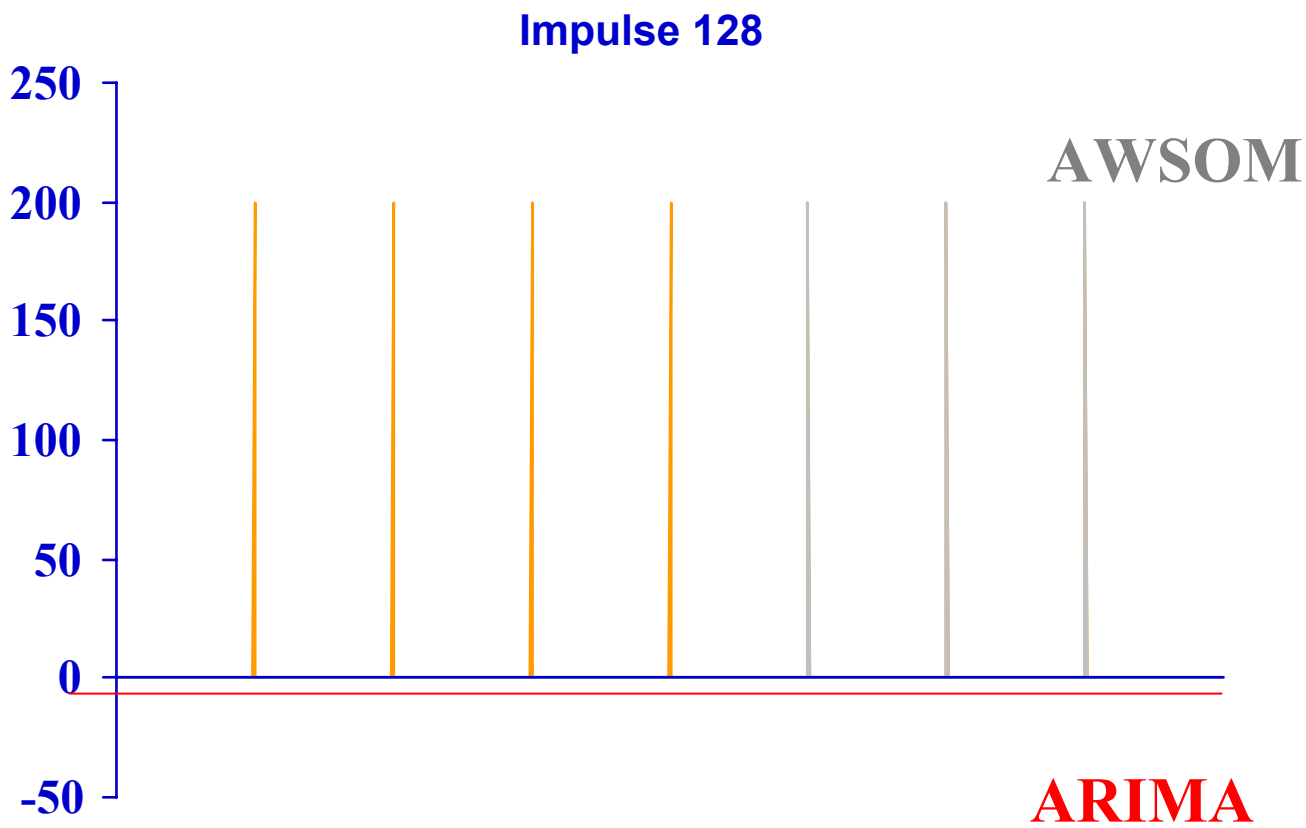
AWSOM: Some Results



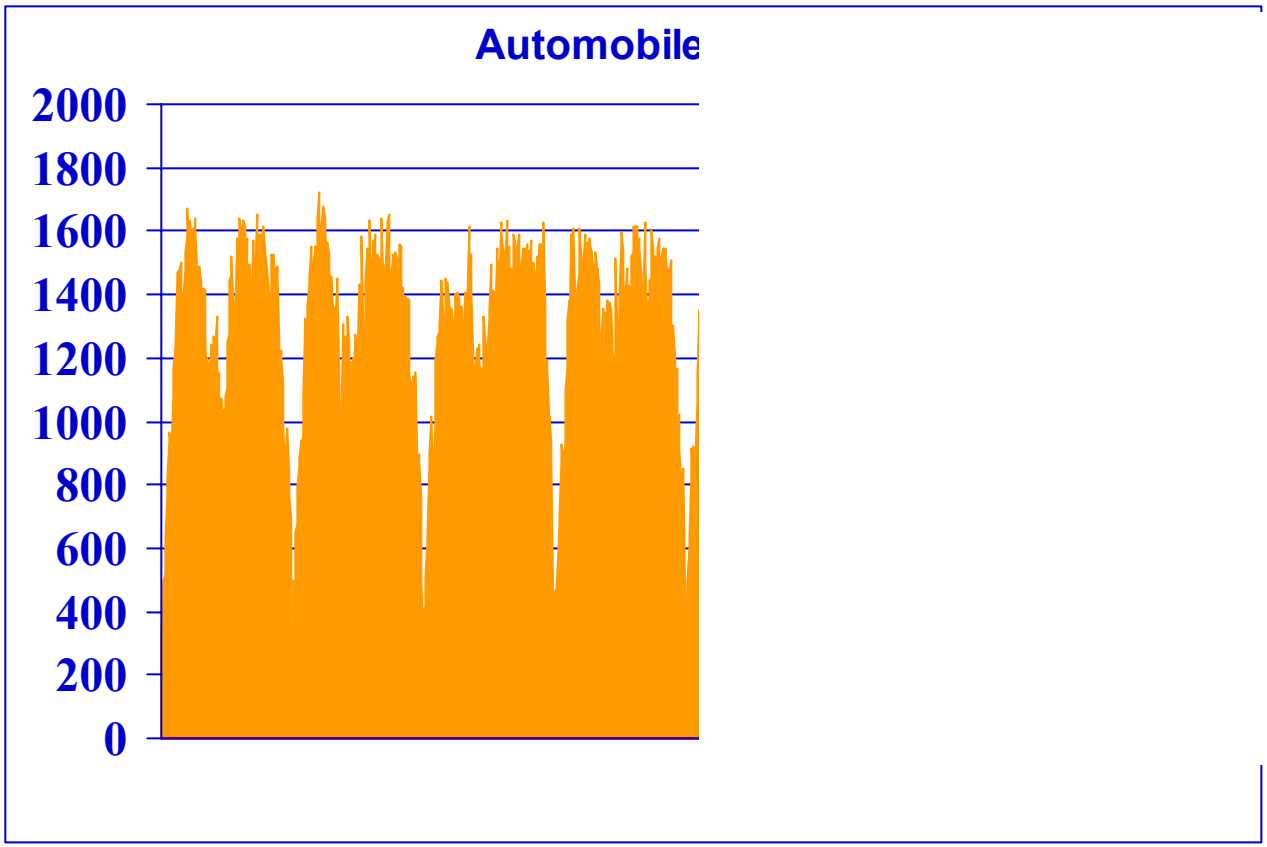
Some Results



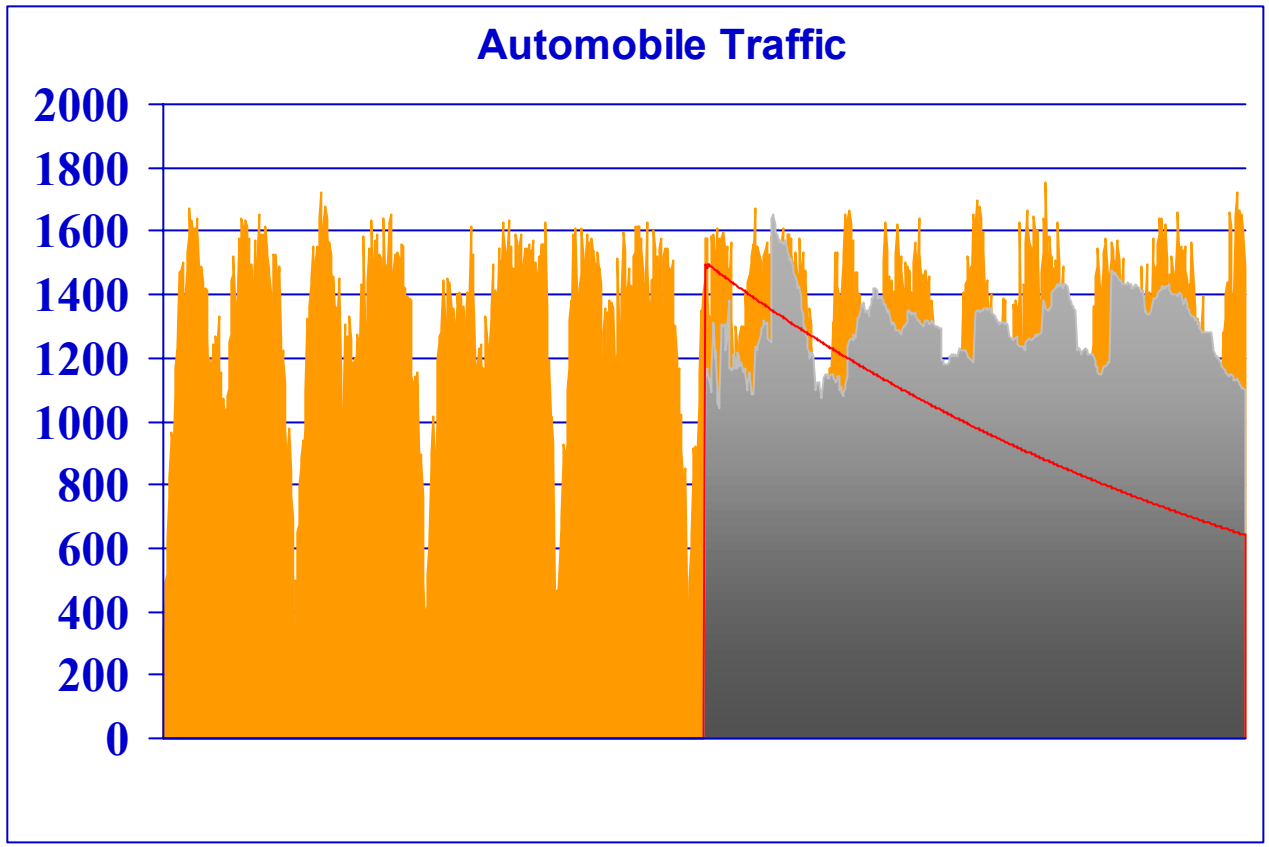
Some Results



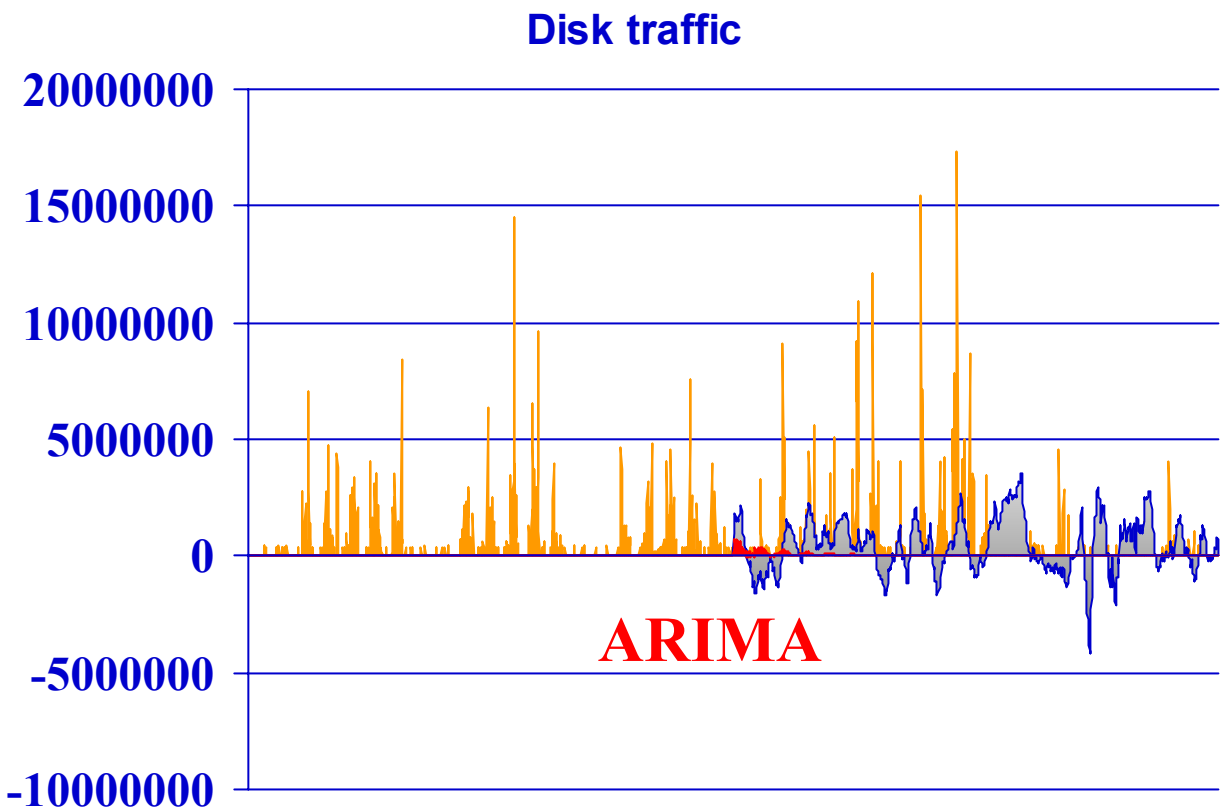
Some Results



Some Results



Some Results



Conclusions

- ANF: Fast approximation for Neighborhood function (time: from **days** to **minutes** - 10x, 100x, 1000x speedup)
- AWSOM: Automatic, 'hands-off' traffic modeling (first of its kind!)